Center for Independent Experts (CIE) Independent Peer Review Report for the 61st Stock Assessment Workshop/Stock Assessment Review Committee (SAW/SARC): Benchmark stock assessments for Atlantic surfclam

Martin Cryer

August 2016

Executive Summary

The benchmark stock assessment for Atlantic surfclam implemented in the Stock Synthesis III software package by NEFSC scientists was reviewed by a panel of four independent reviewers between 19 and 21 July 2016. I consider the assessment to be a credible basis for developing fishery management advice on the fishery. The assessment team and their stock assessment workgroup met all their terms of reference (ToR) except for one related to the stock definition; I do not believe that failing to meet this ToR impacts the quality of information available from the assessment nor the quality of advice that might be developed. The stock assessment was not able to resolve absolute biomass scale with any certainty and almost all information on biomass scale comes from Bayesian priors on survey dredge efficiency, especially for the modified commercial dredge used since 2012. The assessment team therefore focused on relative biomass and trends and these were assessed with much more certainty. New biological reference points (BRPs) based on relative trends were also developed and I believe these are superior to the previous pragmatic BRPs. Base model results, a wide range of sensitivity trials, and supplementary analyses outside the model all suggest that the stock is neither overfished nor experiencing overfishing. Projections, using a wide variety of assumed biomass, catch, and recruitment scenarios, suggest that the stock will not become overfished nor experience overfishing in the next 10 years.

I believe that (even) more diagnostics should be presented for a benchmark assessment, that more use could have been made of Markov chain Monte Carlo (MCMC) runs to develop Bayesian posteriors (uncertainty estimates) and diagnose problems, and that the confidence limits on the trajectory of relative biomass for the whole stock are overstated (Figure 155 in the assessment report is an important graph). But, by and large, this is a good quality stock assessment that has dealt with a number of difficult problems and the lack of contrast caused by the low rate of fishing mortality.

Background

The Northeast Regional Stock Assessment Review Committee (SARC) meeting is a formal meeting of a Panel of stock assessment experts charged with the peer-review of selected stock assessments and models. This report is an independent peer review of benchmark stock assessments for Atlantic surfclam presented at the 61st SARC meeting held at Woods Hole, Massachusetts, 19–21 July 2016. The SARC panel comprised a chairman, Dr Mike Wilberg, and three reviewers appointed by the Center for Independent Experts (CIE), Drs Mike Bell, Coby Needle, and me. This report constitutes my own personal review and perspective of the assessment. It is designed to be read as a stand-alone document, but there are strong overlaps with the Summary Report developed collaboratively with the other members of the review Panel. I agree with all statements made in the Summary Report, and some of the text may be very similar, but this report includes further detail on parts of the stock assessment where I have particular interest or knowledge.

Role in the Review

Most of the necessary background papers for the surfclam assessment were made available on 12 July 2016, but some relating to the SS3 stock assessment software were provided on request a few days later. I read most of these documents before arriving in Falmouth/Woods Hole but, because only a week was available, some I had to read during the review. I had to prioritise my own reading because, as indicated in the Panel's summary report, no particular guidance was provided. I focused primarily on the large stock assessment document itself, annotating the electronic document as I worked though, and read background documents as necessary to aid my understanding. Because I also conducted the SARC-56 review in 2013, my knowledge of the fishery was already reasonable. Electronic copies of the presentations, rapporteur notes, and various additional analyses were made available as they were requested by the Panel.

After meeting briefly with Drs Jim Weinberg and Russ Brown in the early morning of 19 July, the Review Panel worked collaboratively on the Terms of Reference for the Review together with the stock assessment team throughout 19 and 20 July. The Panel met again on 21 July to agree on consensus points and to start drafting the Summary Report. There was strong agreement among the Panel members on almost all aspects of the review and drafting most of the Review Summary Report was straightforward. The Panel Chair allocated Stock Assessment Terms of Reference (ToR) among the Panel members and I tackled ToR5 (stock assessment model) and ToR8 (projections). Toward the end of 22 July, the Chair collated all contributions into a single document. Versions were circulated and comments addressed until an agreed final Summary Report was submitted by the Chair to Dr Jim Weinberg for a check for factual correctness on 28 July. I submitted this individual review on 4 August 2016.

Findings as to whether the work provides a scientifically credible basis for developing fishery management advice

I consider the stock assessment for Atlantic surfclams provides a scientifically credible basis for developing fishery management advice on the fishery. I offer detailed comments in relation to each Term of Reference below.

Findings by Term of Reference (ToR)

1. *Estimate catch from all sources including landings and discards. Map the spatial and temporal distribution of landings, discards, fishing effort, and gross revenue, as appropriate. Characterize the uncertainty in these sources of data.*

I believe the SAW met this Term of Reference.

The commercial surfclam fishery in the US EEZ has operated under a quota system based on individual transferrable quotas, ITQs, since 1990. Hydraulic dredge vessels land their catch in tagged 'industry cages' and logbooks use these same volumetric units. Cage volumes are converted to bushels (which are variable), and subsequently to meat weights (also variable) and numbers at length for use in the stock assessment. The ITQ system also requires logbooks with a spatial resolution of one ten minute square to be completed. Compared with many fisheries, landings data for surfclam are probably accurate and precise. One concern, expressed during the review, was that variability in conversion factors could stem from seasonal (or longer term) changes in surfclam condition, or changes in the size distribution of harvested clams. Both the survey and the quotas are expressed in meats, and fishing practices focused on high clam size and meat yields may obviate some of these potential problems, but I agree with the rest of the Panel that some additional work characterizing and including the uncertainties would be worthwhile.

The surfclam fishery tends to avoid areas where ocean quahog co-occur, resulting in almost no bycatch. Minimum landing sizes were in place from 1982-1990, during which time discards occurred, and were estimated, but discards are now considered negligible (this might need to be checked). Incidental mortality of surfclams caused during dredging is considered low but catches (landings plus incidental mortality) are assumed to be 12% higher than landings. I believe it would be more accurate to estimate incidental mortality as 12% higher than the sum of landings and discards, but this will make little difference to the stock assessment, especially at the very low rate of fishing mortality estimated.

Catch data were treated as exact in the assessment models. Some uncertainties in catches were described and characterized in the stock assessment report, but were not formally quantified and it was assumed that only a single catch history was required. I agree with other members of the Panel that this might be worth further exploration and verification although, given the very low rate of fishing mortality, it is unlikely to affect the conclusions of the assessment or management advice.

The stock assessment report included many figures and tables summarising catch and effort data through space and time, and the patterns were well described both regionally and by ten-minute reporting square. Landings have been stable at ~20,000 t (meats) for many years but there has been a substantial shift to more northern parts of the southern fishery and, more recently to onto Georges Bank where Landings per Unit of Fishing Effort (LPUE) has been very high compared with the southern part of the fishery.

2. *Present the survey data being used in the assessment (e.g., indices of relative or absolute abundance, recruitment, state surveys, age-length data, etc.). Use logbook data to investigate regional changes in LPUE, catch and effort. Characterize the uncertainty and any bias in these sources of data. Evaluate the spatial coverage, precision, and accuracy of the new clam survey.*

I believe the SAW met this Term of Reference.

The assessment working group used data from two surveys, the first from 1982 to 2011 using the research vessel RV *Delaware II* towing a small research dredge (RD), and the second since 2012 using the commercial vessel ESS *Pursuit* towing a modified commercial dredge (MCD). The Panel had substantial discussion of whether it was appropriate to use the MCD survey in the assessment, given the availability of only one year's data for Georges Bank and a different two years' data for southern areas. Without comparative trials between the two platforms or good information on catchability, it is normal practice in many parts of the world for new surveys not to be included in a model until there are at least 3–5 years in the time series. In this case, quite reliable information on

efficiency and selectivity are available, based on depletion experiments using the survey set-up, and this was incorporated as a Bayesian prior on the "q" (catchability) term for MCD in the assessment model. I agree with the Panel's final decision that it was appropriate to include the MCD survey data and its prior, but it is important to note that almost all information on biomass scale in the assessment comes from this prior.

Use of the MCD survey did not result in increased survey precision, as had been expected given its greater fishing power relative to the RD survey, but I agree with the Panel that this was likely because of the need to use incomplete surveys within years and the relatively limited numbers of survey stations. The MCD surveys nominally ascribed to individual years have in fact been split between two years in both northern and southern areas because technical problems limited the spatial coverage possible within individual survey campaigns. It would be much better to complete surveys for given areas or strata within a year or to devised more sophisticated (i.e., spatially structured) methods of fitting surveys within the model.

The survey is based on spatially-stratified random sampling and uses a data "borrowing" strategy to fill gaps where strata or regions have not been surveyed in particular years. This involves replicating, without adjustment, both abundance and composition data from adjacent years for the same stratum or region. Borrowing applies to a greater or lesser extent to all regions and to most years. Some regions have several years of mostly missing data (Southern Virginia from 1999 onwards, Georges Bank during the 1980s), and in some cases borrowing is applied to whole years for particular regions (e.g., Georges Bank in 2005). Otherwise borrowing is patchily distributed by year and stratum. I agree with the Panel and the Working Group that borrowing is not a satisfactory method of data imputation, and is likely to cause smoothing of trends between years, but better approaches have yet to be developed. I agree with the Panel that a more defensible statistical approach to imputation should be developed (e.g., using GLMs/GAMs or geostatistical methods) or a more spatially-structured approach to fitting survey information within the model.

The survey description in the assessment report includes a reference to "nearly random tows" added to surveys to make sure all "areas of interest" received some tows. This was followed up at the review meeting and it transpires that these tows were certainly not random and should not be included in biomass estimates based on assumptions of random site selection. Non-random tows can be useful for monitoring trends and for providing substantial numbers of animals for length or age estimation, but they should not be used for biomass estimation. If the first draw of random tows for a stratum within a survey appears highly clumped or unrepresentative for some reason, it is much better to change the stratification to make sure all "areas of interest" receive some tows, or to enforce a minimum distance between tows within the software (this has the effect of spreading tows throughout strata; it is usually used to overcome spatial correlation). The simplest approach is just to make another draw of random stations.

The Panel noted that the stock assessment treats the survey data for Georges Bank and the southern area separately, even though exactly the same methods, gear and vessels were used. This may be an inefficient use of available information and fitting the two areas simultaneously while sharing some parameters in the assessment (i.e., using a two-stock model) could lead to better precision. Survey results were very well described in the assessment report, and I was particularly impressed with the new work on dredge efficiency; this turns out to have substantial impact on the assessment. The work on the RD presented to SARC-56 in 2013 was a great advance on previous information, even though the results had poor precision, and the work presented here was an even bigger advance. The cooperative trials using repeated tows with commercial and MCD gear to deplete demarked sites were used in a new modelling framework to develop quite precise estimates MCD efficiency. Having attempted similar work for New Zealand dredge fisheries, I am well aware of the problems likely to have been faced and I am, again, very impressed with the work that has been done. As for SARC-56, the estimate of dredge efficiency, with its associated statistical distribution, was used as a Bayesian prior on q for MCD survey(s) in the stock assessment model. There is good reason to be more confident in the prior for the MCD survey than in the prior on the RD survey.

Commercial LPUE values based on logbook data show mostly declining trends in the southern region, a pattern not evident in survey trends. LPUE is usually expected to be more stable than stock biomass so this is probably a result of localized depletion (a combination of fisher behaviour and site selection). Because the fishery impacts only a very small fraction of the overall stock distribution, trends in LPUE are not indicative of trends in the stock as a whole; randomized surveys are better indicators of trends over the whole stock. As for SARC-56 in 2013, these LPUE indices were not included or fitted in the stock assessment model. A more convincing rationale was provided for this review, SARC-61, and the assessment report showed very clearly that LPUE was not at all linearly related to abundance estimated by surveys. Discussions during the review also highlighted the very low fraction of the total stock area impacted by the fishery, so it is not surprising that LPUE within the fishery is a poor index of abundance.

Similarly, declining trends from state surveys carried out in Massachusetts, New York and New Jersey state waters (as Appendix XVIII) but not used in the assessment model. In this case, the indices were not fitted because the modelled population excludes these coastal waters. It would be worth reconsidering that decision for future assessments, especially if a more spatially structured approach is contemplated.

3. *Determine the extent and relative quality of benthic habitat for Atlantic surfclam in the Georges Bank ecosystem to refine estimates of stock size based on swept area calculations.*

I believe this ToR was adequately met in relation to habitat extent (for the purposes of this assessment) but, for good reason, little work was presented on habitat quality.

Based on the finding that 14% of attempted stations were untrawlable, the assessment team concluded that 14% of ground is potentially poor clam habitat. I agree that this is a reasonable inference based on the limited information available, and it is only slightly higher than the 12% figure used in the assessment. Discussion revealed that new, much more comprehensive analyses based on a variety of new data would be conducted soon, and these would provide much more defensible estimates than the existing ones. Given the very low estimated rate of fishing mortality and the low fraction of the total area of the stock impacted by the fishery each year, spending a lot of time on another slightly better approximate estimate of the extent and/or quality of available habitat when a much better estimate will be available quite soon would not have been a good use of the assessment team's resources or the workgroup's time.

4. *Quantify changes in the depth distribution of Atlantic surfclam over time. Review changes over time in Atlantic surfclam biological parameters such as length, width, and growth.*

I believe the SAW met this Term of Reference.

The assessment report included strong evidence that surfclams are shifting into deeper waters, at least in the southern part of its range, probably in response to increasing water temperatures. Analysis also showed increasing spatial overlap between Atlantic surfclam and ocean quahog, and this may become problematic for fishers attempting to maintain relatively "clean" catches of one or the other species. The working group also demonstrated changes in regional growth patterns over time, which may be attributable to increased ocean temperature. I agree with the rest of the Panel that it will be important to examine the drivers of these changes and the implications for stock productivity and the assessment model(s).

5. *Estimate annual fishing mortality, recruitment and stock biomass (both total and spawning stock) for the time series (integrating results from TOR 3, as appropriate) and estimate their*

uncertainty. Include a historical retrospective analysis to allow a comparison with previous assessment results and previous projections.

I believe this ToR was met to the extent possible and necessary (given the difficulties in estimating biomass scale in the model using the information available).

The primary assessment approach used integrated statistical catch-at-age models implemented in Stock Synthesis III (SS3). I agree that SS3 was an appropriate choice of platform for this assessment (as it was for the first time in SARC-56 in 2013) because it allowed the integration within a single model of biomass indices, conditional age compositions, size compositions, and other information from various sources. Other integrated stock assessment packages could be used, of course, but there is value in focusing institutional capacity around relatively few packages and having a "critical mass" of users who can share ideas and scripts and mentor new users.

Georges Bank was assessed in a separate model from the other area (collectively called the Southern Area) and the results were combined to provide management advice for the assumed stock. The panel agrees that this is probably the best way to conduct the stock assessment modelling, although it does introduce some complications when developing advice for the assumed single stock and it may be appropriate to consider whether consistency between assessment area and management area would be achievable. The scale of absolute abundance was uncertain in both models, but the trend in relative biomass was shown by the assessment team to be much more certain. A very wide range of sensitivity analyses were conducted, and detailed results were provided to the panel as well as the numerous plots and tables in the assessment report.

Based mostly on my experience in New Zealand where substantially more detail is usually reported, especially for important assessments or those targeted at external review, I think insufficient detail was provided in the report on the process that led to the working group accepting base models for each area. The base models had labels of BASE7 and GBKBASE7 suggesting that several iterations had occurred during development, but none of this was described in the report. This can be important because it is during that development process that many important modeling choices are made relating to structure (e.g., separate models for Southern and Georges Bank) or data (e.g., exclusion of State surveys and fishery LPUE, and the choice of conditional age at length). This is an area where the stock assessment report could be much more compelling, leading the reader through a structured progression of modeling choices supported by fits, residuals, and other diagnostics. In New Zealand, I am accustomed to seeing comprehensive tables of objective functions, weights, contributions to likelihoods and SDNRs for the main data sets and the main parameter estimates. These can be dense tables, but they provide valuable comparisons among models and can be used to show a logical progression in model choices.

I also agree with the rest of the Panel that more detail of the methods used should have been provided in the assessment report, including the actual equations used (probably in an appendix). SS3 is only one of several integrated packages, and it is not reasonable to expect a reader to be able to interpret the meaning of SS3 "settings" or switches, let alone SS3's default settings. Additional documentation of SS3 was provided to the Panel on request, including the latest version of the use manual, which I found extremely useful, but I think it would be better to include key equations in the actual assessment report. This would make an already large report (470 pages) larger still, but the increasing use of electronic files instead of paper reports means that this is less of an issue than it once was. The aim should be to provide a report that is sufficiently detailed to enable a new analyst or another organization to closely replicate the assessment model with little input from the current analyst, even if using different software.

I agree with the rest of the Panel that the following assumptions or approaches used in the modelling were appropriate:

- In contrast with the assessment reported to SARC-56 in 2013, the assessment team in 2016 gave higher priority to fitting the survey indices than to fitting the length and age composition data (because trends in biomass are particularly important for this assessment);
- Excluding the LPUE data in the assessment model (because of the availability of fishery independent data, the small proportion of the stock area fished, and the demonstrably poor relationship between LPUE and survey indices of wider abundance);
- The use, for this assessment (but not as a generality) of "borrowing" tows from adjacent years to fill gaps in some strata within the survey time series (because other choices were limited and it was considered the impact was likely to be minor, although more statistically robust approaches either outside or inside the assessment model should be considered in future);
- The use of a wide range of sensitivity runs to understand the impact of modelling choices;
- The focus on trends in stock size and fishing mortality rates (because trends were estimated much more reliably than absolute biomass and because almost all of the information on biomass scale came from priors on catchability); and
- The use of supplementary analyses outside the model to corroborate the results of the SS3 assessment.

All members of the Panel brought up issues during discussions; the following are those that I think are particularly important and worthy of further work:

- The priors for catchability were estimated differently for the old and new surveys; in general, I think the depletion methodology for MCD surveys was better than the comparative method used for RD surveys, but the depletion implementation may be biased if the precise location of the dredge on the bottom is not known;
- Catchability for the survey in the model may not be exactly the same as assumed by the prior; in particular, if significant parts of the fishery are not included in the survey area (e.g., Nantucket Shoals). This would become much more important, of course, if fishing mortality were to increase or absolute estimates of abundance were to be required from the model rather than relative estimates;
- The assumed dome-shaped selectivity patterns for the survey were based on gear selectivity experiments and are not identical to the way selectivity is defined in the model; the commercial selectivity pattern was not reflective of the experimentally-estimated gear selectivity because large individuals are not fully selected by the fishery. The cause of this is probably fishery behavior, especially of the choice of areas to be fished;
- There were conflicts among priors and some data sets for both models, but especially that for the Southern Area. Fits to some early length and age distributions were particularly poor. This is a common problem in integrated stock assessments but may be indicative of structural problems that should be explored in subsequent stock assessments. One possible cause is un-modelled heterogeneity (in space or time or both) in growth, recruitment, or mortality, and a more structured assessment model might be trialed to address them;
- My personal belief is that Markov chain Monte Carlo (MCMC) runs should be used to characterize uncertainty for all purportedly Bayesian stock assessment models used to develop management advice, especially risk-based advice. MCMCs can be particularly useful for describing uncertainty in derived quantities like $SSB/SSB_{Threshold}$, and can also provide powerful diagnostic tools, especially to identify poorly-determined parameters and structural problems. MCMCs naturally maintain correlations between parameters and exploring the form of these correlations can be informative. Occasionally, MCMC results are somewhat different from MPD (actually penalized maximum likelihood) estimates, and this can also give important diagnostic information;
- The Panel was very grateful for the single MCMC chain run by the assessment team for the Southern Area base case during the review. The trace shown to the Panel was well-formed and appeared to have converged, although no formal tests were shown. It gave a much better estimate of uncertainty for a key parameter (q for MCD surveys) than the analytical output provided. I think it would have been extremely useful if MCMC runs had been done for at least for the two base cases and reported in substantial detail, including traces and

convergence tests for all estimated parameters. SS3 and the R-package R4SS provide ready-built code for developing some of this output and I believe it should be standard for all Bayesian assessments;

- The rate of natural mortality, M, is assumed and fixed at 0.15 within the model. This is slightly higher than the available estimates of Z and lower assumed values for M may have led, as they did in SARC-56 in 2013, to different results (although one sensitivity model that estimated M during fitting resulted in a similar level). I think it would have been useful to run a sensitivity model with a lower assumed M, perhaps M = 0.10, alongside the many others. This has been a recurring theme (SARC-49, SARC-56, and SARC-61); and
- Catch-curve analysis in unfished and fished areas can provide estimates of M and Z, respectively. If these are contemplated for the future, better methods than simple or weighted regression should be used (e.g., the Chapman-Robson estimator which was shown by Dunn et al. (2002) to be superior to regression methods or, better still, the model-based estimation method described by Millar (2014).

Despite the uncertainty over biomass scale, I agree with the rest of the Panel that the two assessment models are credible and strongly indicate that stock biomass is above the target, and fishing mortality is low for both areas; the stock is neither overfished nor experiencing overfishing. This conclusion is driven by the robustly-determined trends in relative biomass within the two models and corroborated by the following supplementary analyses presented to the Panel:

- Z estimates from catch curves were about the same as the assumed M (indicating that, although outdated methods were used, the value of F must be low);
- The proportion of the stock area fished each year (i.e., aggregate area swept divided by the total stock area) was about the same as the estimated F;
- Exploitation rates estimated by dividing catch by swept area biomass estimates from surveys were similar to estimated F; and
- The average age of individuals on Georges Bank (unfished for about 20 years) was about the same as the average age in the southern region (where fishing has been continuous).

Somewhat strangely, these corroboratory analyses were held back to the end of the discussion at this review, and some were conducted overnight by the assessment team, whereas several such analyses were presented as "context" in the very first presentation to the SARC-56 Panel in 2013.

One outstanding concern I have with the presentation and communication of the assessment (not with the assessment itself) is with the estimated confidence limits for the trend in relative biomass for the entire stock. I believe the estimated confidence limits presented in Figure 155 of the assessment report and duplicated in the assessment summary report are much too broad. I don't think this figure reflects the much higher precision with which relative biomass trends are estimated for the Southern Area (especially, see Figure 129) and the Georges Bank Area (see Figure 146, see also where I combine these three figures below in my Figure 1). I brought this up at the review meeting and there was some discussion. The assessment team later kindly checked that their code generated the confidence limits that they intended, using the pooled CV method. On the face of it, this seems like a defensible approach given that the two separate model results had similar means, but I was sufficiently concerned to conduct my own quick calculations after the meeting.

I think it is worth describing my calculations because I believe that, based on the information I have at hand, they reinforce my concerns. I extracted what I believe to be the estimated derived quantities of interest (the biomass ratio and its estimated standard deviation for the two base cases) from the SS3 output file named *Report.sso* in each of the two directories named BASE7 and GBKBASE7. These are tabulated in my Table 1. To simplify my calculations, and because the confidence limits in Figures 129 and 146 were almost symmetrical, I chose to assume that the errors were normally distributed. Also for simplicity, I assumed a 50:50 distribution of unfished biomass between the two areas, for which separate estimates of the biomass ratio were available. This meant that, in my calculations, the trend in the biomass ratio for the whole stock was simply the mean of the two biomass ratios; it looks

very much like the trends presented in the assessment document. The trends and approximate confidence limits (± 2 * SD) that I was able to plot for the Southern and Georges Bank areas also look very similar to those presented in the assessment report, suggesting I had indeed extracted the correct output data from the *Report.sso* file. I estimated the approximate confidence interval for the combined trajectory as ± 2 * sqrt(var(Bratio_S)+var(Bratio_N)) where the variances were the square of the estimated standard deviations in the *Report.sso* files. I got very similar results using the pooled CV approach when I calculated the CV for each year as the standard deviation divided by the biomass ratio from the *Report.sso* files.

The approximate confidence intervals calculated this way are much narrower than those shown in Figure 155 of the assessment report (see my Figure 2), but I think they much more accurately reflect what should be the confidence limits for the relative biomass trends across the whole stock, given the confidence limits of the trends in the two constituent areas. I think the use of lognormal errors or different distributions of unfished biomass between areas would make very little difference to this result compared with the very large difference between my approximate confidence limits and those presented in the assessment report and the assessment summary report. I may have misinterpreted the *Report.sso* files or how the assessment team generated their confidence limits, but I do not think so.

I have not been able to replicate the very wide confidence intervals shown in the assessment report and I cannot say how they have come about, but I remain skeptical of them, and I believe they are misleading. This is unfortunate because I think Figure 155 (especially when duplicated in the assessment summary report without the underlying information) tells a story of very poor knowledge of trends in relative biomass for this stock; from all the other material presented and discussed I do not believe this to be the case. This issue does not in any way affect the high quality of the stock assessment, it is simply the impression given by the graphic in the assessment summary report.

Table 1: Trajectories of the biomass ratio ($B_{EST}/B_{THRESHOLD}$) and its estimated standard deviation (SD) from the Southern Area (S) and Georges Bank (N) from respective *Report.sso* files.

YEAR	Value_S	SD_S	Value_N	SD_N
Bratio_1985	5.17286	0.44609	3.96510	0.42174
Bratio_1986	4.78418	0.41294	3.84352	0.41517
Bratio_1987	4.49329	0.38588	3.81601	0.41954
Bratio_1988	4.31740	0.37519	3.82430	0.42537
Bratio_1989	4.28362	0.38029	3.86461	0.43957
Bratio_1990	4.40585	0.41629	3.72994	0.43520
Bratio_1991	4.54261	0.44811	3.54802	0.41998
Bratio_1992	5.08163	0.51779	3.60100	0.44702
Bratio_1993	6.66669	0.62145	3.68840	0.47282
Bratio_1994	7.39215	0.69025	4.90162	0.67069
Bratio_1995	6.82979	0.60392	5.91277	0.81887
Bratio_1996	6.19800	0.54108	6.21950	0.85857
Bratio_1997	5.58238	0.48387	6.20741	0.85454
Bratio_1998	5.03785	0.43580	6.32292	0.88439
Bratio_1999	4.63505	0.39887	6.72128	0.93301
Bratio_2000	4.31077	0.36967	6.28218	0.84538
Bratio_2001	3.91428	0.33598	5.51780	0.72653
Bratio_2002	3.47737	0.30126	4.83485	0.62291
Bratio_2003	3.08132	0.27034	4.34458	0.54237
Bratio_2004	2.74434	0.24475	3.91961	0.47190
Bratio_2005	2.49382	0.22664	3.67639	0.42657
Bratio_2006	2.34609	0.21706	3.49640	0.39343
Bratio_2007	2.28633	0.21405	3.43983	0.39006
Bratio_2008	2.34523	0.23753	3.17390	0.35163
Bratio_2009	2.57592	0.28362	2.90171	0.31235
Bratio_2010	2.89381	0.30232	2.67294	0.28470
Bratio_2011	3.20955	0.35431	2.37478	0.24802
Bratio_2012	3.26130	0.42224	2.14958	0.21991
Bratio_2013	3.24519	0.43307	2.00225	0.21760
Bratio_2014	3.24950	0.48526	2.00082	0.30417
Bratio_2015	3.19515	0.50100	2.01744	0.36489

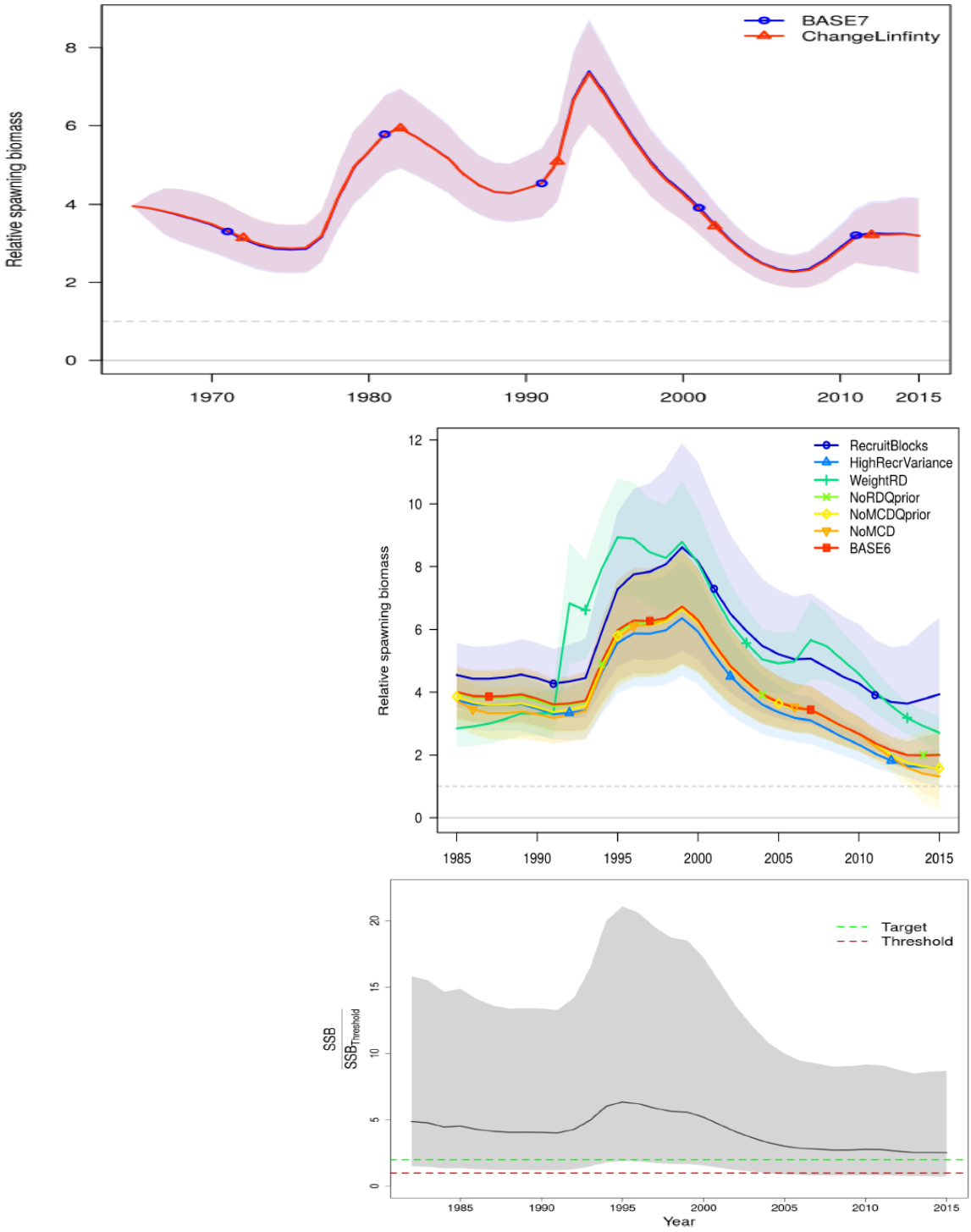

Figure 1: Relative biomass trajectories ($SSB_{EST}/SSB_{UNFISHED}$) for the Southern Area (top, after Fig. 129 of the assessment report), Georges Bank (middle, after Fig 146), and for the combined stock (bottom, after Fig 155).

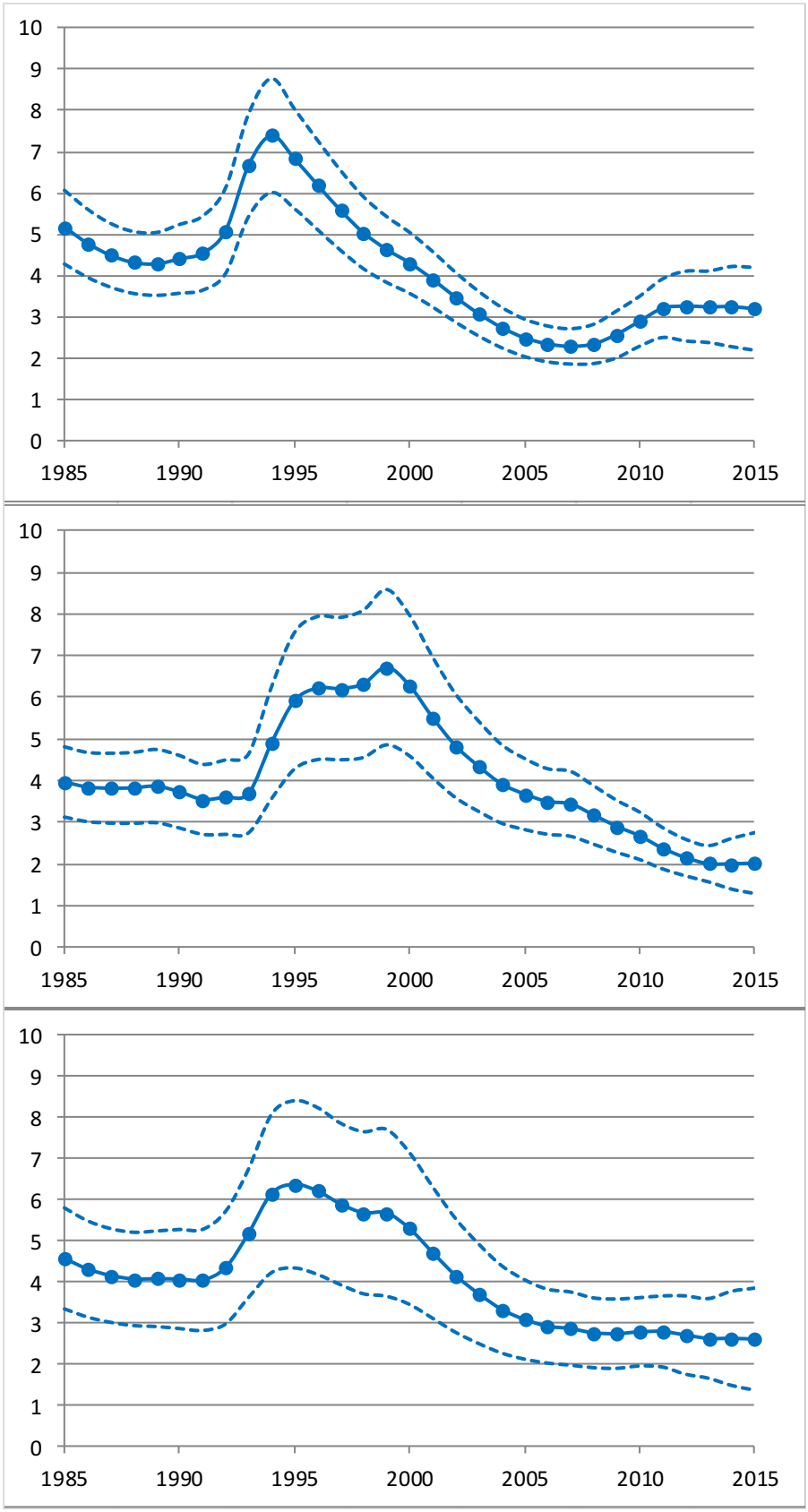

Figure 2: Relative biomass trajectories ($SSB_{EST}/SSB_{UNFISHED}$) for the Southern Area (top) and Georges Bank (middle, both using SS3 output files *Report.sso*), and for the combined stock (bottom, with approximate normal confidence limits estimated as twice the square root of the summed variances for the Southern and Georges Bank areas.

6. *State the existing stock status definitions for "overfished" and "overfishing". Then update or redefine biological reference points (BRPs; point estimates or proxies for B_{MSY}, $B_{THRESHOLD}$, F_{MSY} and MSY) and provide estimates of their uncertainty. If analytic model-based estimates are unavailable, consider recommending alternative measurable proxies for BRPs. Comment on the scientific adequacy of existing BRPs and the "new" (i.e., updated, redefined, or alternative) BRPs, particularly as they relate to stock assumptions.*

I believe that this ToR was met.

I agree with the rest of the Panel that the proposed new biological reference points (BRPs) based on relative stock status should be accepted. The new BRPs are theoretically more defensible yet are relatively simple and appear robust to the uncertainty of population scale (a major uncertainty in this assessment). I believe the new BRPs are suitable for providing advice on catch limits.

The Panel noted that the fishing mortality threshold calculation uses an estimate of F_{MSY}. This value was derived from a simulation study which was not discussed in much detail by the Panel but was described in the documentation. I agree with the rest of the Panel that the use of the estimated value of F_{MSY} to estimate the BRP for the overfishing threshold was reasonable but that the estimated F_{MSY} value should not (yet) be used as a stand-alone reference point.

The use of the new BRPs should lead to status determinations being more stable in future assessments, even if the population scale estimated by the assessment changes.

7. *Evaluate stock status with respect to the existing model (from previous peer reviewed accepted assessment) and with respect to any new model or models developed for this peer review.*
    a. *When working with the existing model, update it with new data and evaluate stock status (overfished and overfishing) with respect to the existing BRP estimates.*
    b. *Then use the newly proposed model and evaluate stock status with respect to "new" BRPs and their estimates (from TOR-5).*

I believe this ToR was met.

All of the analyses conducted indicated that the stock was not overfished nor was it experiencing overfishing. This result was consistent across stock areas and a wide variety of sensitivity runs. Given the large uncertainty in the estimates of population scale, there remained a small probability that the stock was overfished in the most recent years. However, I agree with the other members of the Panel that improved approaches to stock modelling (especially sharing information on survey efficiency between areas) and characterizing uncertainty (especially, in my opinion, using MCMC runs) will demonstrate that an overfished stock status is not consistent with the data.

I think, however, that it is quite likely that the very small proportion of the stock within the Southern Area that has been repeatedly fished could be overfished and, perhaps, be experiencing overfishing. Given the current stock definition (and any conceivable stock definition for management purposes) and the wide spread of recruiting individuals across the stock area, this should not be interpreted as a biological problem, but it could indicate some loss of productivity within the fished areas. Wider spreading of fishing effort could increase productivity, but this would be feasible only if the costs of fishing other areas are not substantially greater.

8. *Develop approaches and apply them to conduct stock projections.*
    a. *Provide numerical annual projections (five years) and the statistical distribution (e.g., probability density function) of the OFL (overfishing level) (see Appendix to the*

SAW TORs). Consider cases using nominal as well as potential levels of uncertainty in the model. Each projection should estimate and report annual probabilities of exceeding threshold BRPs for F, and probabilities of falling below threshold BRPs for biomass. Use a sensitivity analysis approach in which a range of assumptions about the most important uncertainties in the assessment are considered (e.g., terminal year abundance, variability in recruitment).

b. *Comment on which projections seem most realistic. Consider the major uncertainties in the assessment as well as sensitivity of the projections to various assumptions.*
c. *Describe this stock's vulnerability (see "Appendix to the SAW TORs") to becoming overfished, and how this could affect the choice of ABC.*

I believe that this ToR was largely met, subject to the treatment of predicted recruitment within projections.

The projections provided in the report suggested that the population is unlikely to become overfished and that overfishing is unlikely to occur by 2025 using a wide range of possible biomass scales and assumed catches. However, these projections assumed average recruitment estimated over the whole modelled history of the fishery, ignored temporal autocorrelation, and included stochasticity only in annual recruitment. I agree with the rest of the panel that it would be better to take a more holistic view of uncertainty and include more of the estimated parameter uncertainties from the assessment models as well as temporal autocorrelation in recruitment.

I also think it should be standard practice to conduct projections using different historical time periods as a basis for predicting future recruitment. For example, it is common practice for projections using recent average recruitment to be conducted (based on the assumption that the recent past is more likely to be similar to the near future) and for these to be contrasted with projections assuming long-run average recruitment. Alternatively, or in addition, high and low recruitment scenarios can be modelled to test the robustness of estimates of risk.

During the meeting, additional projections using recent average recruitment or the lowest historical 10-year period were requested by the Panel. The assessment team had some difficulty implementing either within SS3, and instead ran projections sampling randomly (i.e., ignoring any temporal autocorrelation) from the ten years with the lowest recruitment estimates from the entire time series. This is a very conservative approach but these projections confirmed the assessment team's conclusion that the population is unlikely to become overfished, and that overfishing is unlikely to occur by 2025 using a wide range of assumed catches, even if recruitment is as low as it ever has been.

9. *Evaluate the validity of the current stock definition. Determine whether current stock definitions may mask reductions in sustainable catch on regional spatial scales. Make a recommendation about whether there is a need to modify the current stock definition.*

I agree with other members of the Panel that this ToR was <u>not</u> met and that not meeting this ToR did not impact the quality of information from the assessment.

The assessment working group did not reach a consensus on whether the Atlantic surfclam resource should be considered as one unit stock throughout the species' range in US waters or if regional stocks should be recognized. SARC/SAW-56 had previously described the arguments for and against splitting surfclam into north (Georges Bank) and south (Southern Virginia to Southern New England) stocks. The arguments for or against hinged on whether emphasis is placed on regional differences in biology and meaningful definition of MSY (favoring two stocks), or allowing flexibility for the fishing industry and avoidance of unnecessary management intervention (favouring one stock). It was also recognized that there are clinal (i.e. gradual) differences within the resource, and other stock divisions that might be equally valid. Similar arguments could be used to defend (or not) the split

between state and federal waters in the assessment. Given no further basis for resolving these issues since they were previously addressed in 2013, the working group chose not to address these further than a simple re-statement in the SARC-61 stock assessment report. All members of the working group agree that stock definitions are unlikely to affect management, yield, or biological risk in the near term as long as fishing mortality rates remain low, and overall abundance and biomass are relatively high in both the northern and southern areas.

I agree that a lack of progress with this ToR has not affected the outcomes of the assessment or the quality of management advice, and I support the pragmatic decision to combine stock assessment outputs across northern and southern areas after modelling the populations separately. However, if fishing mortality increases substantially, or some portion of the stock declines substantially, then the current stock definition has the potential to mask conditions in the affected area and lead to reduced yield and biomass. I am sure that assessment reports and advice will always include detail on declines in biomass that particularly or only affect some parts of the stock, but the decision-making process may not always be able to incorporate this advice if the stock definition is for a single stock. I am not sufficiently familiar with the US decision-making system to know how big this risk might be.

The spatial scales of stock definition considered by the working group are likely larger than the scale at which depletion occurs at current fishing intensity, as well as the scales at which stock dynamic processes occur. Although management and full stock assessment at these small spatial scales is unlikely to be either practicable or desirable, I agree with the Panel's recommendation that survey and commercial catch data should be explored to determine spatial scales of variability and examine the implications of these for assessing and managing at larger spatial scales.

10. *Review, evaluate and report on the status of the SARC and Working Group research recommendations listed in most recent SARC reviewed assessment and review panel reports. Identify new research recommendations.*

I believe this ToR was met.

The assessment working group provided updated information on work that had been completed on the five Research Recommendations from the 2013 assessment and provided three new research recommendations. I agree with the rest of the Panel that the three new research recommendations would provide useful information for assessment and management. In addition, the Panel spent some time discussing additional research that could improve the assessment of Atlantic surfclam:

- A key recommendation from the Panel is to carefully consider any new changes made to the NEFSC clam survey. This survey provides the primary estimate of scale of the population, and its continuation will be important for stabilizing assessment estimates of biomass and fishing mortality rates. Further changes in the survey could introduce problems in future assessments because changes in efficiency of the dredge (particularly to a less efficient one) would increase uncertainty in the population scale. Therefore, I strongly support the Panel's recommendation that the current (MCD) gear and vessel is retained within any modified survey design.
- If the spatial scope of the survey is expanded to include new regions, like Nantucket Shoals, it will be necessary to carefully consider how to treat the area of the stock for the previous survey. If there is a mismatch in scale, it could cause problems in interpreting changes in indices from one survey to another that surveyed a different spatial scale. Also, the survey has not been implemented as planned in the most recent years, with some areas not being fully surveyed in the intended year. This has caused problems for developing an index of abundance because of the need to aggregate data over years. If the survey for a region cannot be completed in the intended year, one way to improve estimates of the index would be to ensure that some of the strata are sampled in both years so that a "year effect" can be estimated for the survey index using a model-based approach.

- The assessment model assumed that an additional 12% of the landings were killed during fishing activities. The panel thought it likely that this value was size-dependent and may have changed with the prosecution of the fishery. This value should be reexamined using observed data, and re-estimated if necessary. Additionally, the value of incidental mortality may be able to be evaluated during a study that also estimates improved size-meat weight conversions.
- The efficiency of the dredge is very important for setting the scale of the assessment model and interpreting results from the survey and comparing them to fishery removals. I agree with the Panel's recommendation that depletion experiments to estimate dredge efficiency continue and are improved by better estimating the location of the dredge on the bottom.
- I strongly support the Panel's recommendation to explore whether the structural decisions in the assessment model are leading to the conflicts among the data sources. Some of the most important structural decisions are assuming known selectivity for the survey (and its dome-shape), informative priors for survey gear efficiency, and the variation in size-at-age.
- I agree with the Panel's recommendation to explore methods to simultaneously estimate M within the assessment using data on shells from recently dead individuals (e.g., Wilberg et al. 2011).
- I agree with the Panel's recommendation to maintain the institutional capacity and support for the use of age- and length-based integrated models, including SS3. Focusing on tried-and-tested, well-understood software maximizes consistency and productivity and provides for a critical mass of users who can share ideas and mentor new users.
- I agree with the Panel's recommendation to investigate spatial scales of variability in survey and commercial catch data as it may be useful in improving the design of the survey or in developing regions for assessment or management.
- I agree with the Panel's recommendation to develop model-based or statistical estimators to deal with missing data in surveys, because these should be more defensible than the current data borrowing approach.
- I agree with the Panel's recommendation that discard data from the new observer programme be considered for incorporation in the next assessment, if significant evidence of non-zero discarding exists.
- I agree with the Panel's suggestion that the next assessment considers whether a combined state-federal assessment would be more appropriate, if it is possible to do so.

General comments on the surfclam assessment

I support the use of integrated stock assessment models like the one for Atlantic surfclam when multiple sources of data are available; fitting a model readily identifies conflicts among data sets and forces the assessment team to consider those conflicts explicitly rather than obliquely (or covertly) outside a model. But integrated models are complex and require significant documentation to communicate and to realize these benefits. My first reaction to the surfclam stock assessment at SARC-56 in 2013 was that the report was very large but curiously uninformative on some key aspects of the modeling, especially the key choices made during model development and choice of the base case. I had some similar misgivings this year but, in contrast to SARC-56, the assessment team conducted extensive sensitivity analysis around different potential states of nature (around M, selectivity, recruitment, etc.) and provided analysis to support the exclusion of data sets that I previously thought could potentially be useful. The building of the report using LaTeX with hyperlinked figures and tables made the document easy to read electronically (for a large document). Even though the assessment report is already large, I think more should be added for a review like this, especially the actual equations used (for non-SS3 users) and more diagnostic tables and plots. I enjoyed the assessment team's presentations to the panel, and found them uniformly informative. I was particularly impressed with the team's openness and willingness to conduct additional supporting analyses.

Despite the difficulties faced by the assessment team, especially in relation to biomass scale, they have assembled some powerful new data and their conclusion that this is a large resource that is very lightly exploited appears robust to most uncertainties.

Comments on the NMFS review process

I found the SARC meeting well organized, and most of the necessary background reading for both stock assessments was provided at least a few days in advance. I think it would help if the documents could be provided sooner (say 10 days in advance of the meeting) but I recognize that this is sometimes difficult. Senior NEFSC staff gave the Panel a good introduction to the assessment and process for the review on the first morning, and were always available to provide guidance on process during the three days when the Panel was in session. Excellent rapporteurs and skilled assistance with modifying the summary document were made available.

I thought the open meeting format was good for providing transparency to stakeholders and expert opinion for the panel as and when required. The phone link did not always seem to work as anticipated and was sometimes a little obtrusive, but this is a small price to pay for the additional transparency and accessibility. The meeting was conducted in excellent humour, and the assessment team was open to suggestions and questions, and often willing to undertake additional analyses and model runs for the benefit of the Panel.

As I indicate above, the stock assessment report provided did not contain the detail I think is required for full peer review, especially for a complex model and this may be the norm for NEFSC or even in the US. I was not presented with the wide and informative array of diagnostic plots and tables that I am accustomed to seeing routinely in New Zealand, especially for Bayesian models. I recommend tables of key parameter estimates (and derived quantities), contributions to likelihoods, SDNRs, priors, bounds, etc. as a means to show the rationale for modeling choices. For a purportedly Bayesian model, I would always expect to see MCMC traces for at least the key parameters (and any that behave badly, mixing poorly or running up against bounds) and convergence diagnostics like running means and cumulative plots of two or more independent chains for key parameters. Plots of the distributions of posteriors and comparisons with priors and MPDs are useful, and should be routinely available from the standard software packages (many are from SS3 using R4SS). These diagnostics provide a solid rationale for key modeling choices like data weighting or fixing M or selectivity. One recent example of a report with such detail (a stock assessment for New Zealand hoki by McKenzie 2016) can be found at:

http://www.mpi.govt.nz/document-vault/11623

This is, of course, not the only report with such detail; I chose it because it shows an assessment with two putative stocks and very detailed assessment of MCMC diagnostics.

Appendix 1: Bibliography of materials provided for review

Working Papers

Working Group, Stock Assessment Workshop (SAW 61) (2016). Stock Assessment Report of Atlantic Surfclam. SAW/SARC 61. July 19-21, 2016, NOAA Fisheries, Northeast Fisheries Science Center. Woods Hole, MA. 474 p.

Working Group, Stock Assessment Workshop (SAW 61) (2016). Stock Assessment Summary Report of Atlantic Surfclam. SAW/SARC 61. July 19-21, 2016, NOAA Fisheries, Northeast Fisheries Science Center. Woods Hole, MA. 12 p.

Background Papers

Methot RD (2015). User Manual for Stock Synthesis — Model Version 3.24s. NOAA Fisheries Toolbox. 152 p.

Methot RD, Wetzel C. (2013). Appendix A: Technical Description of the Stock Synthesis assessment program. NOAA Fisheries Toolbox. 37 p.

Munroe DM et al. (2016). Fishing and bottom water temperature as drivers of change in maximum shell length in Atlantic surfclams (*Spisula solidissima*). *Estuarine, Coastal and Shelf Science 170*: 112–122.

NOAA (1999). Essential Fish Habitat Source Document: Atlantic Surfclam, *Spisula solidissima*, Life History and Habitat Characteristics. NOAA Technical Memorandum TM F/NE-142. 22p.

Northeast Fisheries Science Center (2013). 56th Northeast Regional Stock Assessment Workshop (56th SAW) Assessment Report. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 13-10; 868 p.

Northeast Fisheries Science Center (2010). 49th Northeast Regional Stock Assessment Workshop (49th SAW) Assessment Report. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 10-03; 383 p.

Serchuk FM and Murawski SA (1997). The Offshore Molluscan Resources of the Northeastern Coast of the United States: Surfclams, Ocean Quahogs, and Sea Scallops. NOAA Technical Report NMFS 127: 18 p.

Weinberg JR (2005). Bathymetric shift in the distribution of Atlantic surfclams: response to warmer ocean temperature. *ICES Journal of Marine Science 62*: 1444–1453.

Zhang X et al. (2015). Modeling larval connectivity of the Atlantic surfclams within the Middle Atlantic Bight: Model development, larval dispersal and metapopulation connectivity. *Estuarine, Coastal and Shelf Science 153*: 38–53.

Zhang X et al. 2016. Atlantic surfclam connectivity within the Middle Atlantic Bight: Mechanisms underlying variation in larval transport and settlement. *Estuarine, Coastal and Shelf Science 173*: 65–78.

Powerpoint presentations

Working Group, Atlantic Surfclam (2016). Atlantic Surfclam Assessment 2016. PowerPoint presentation. 78 slides.

Jacobson L (2016). Surfclam Swept Area Data. PowerPoint presentation. 7 slides.

Jacobson L (2016). Overfishing Limits Based on Trends. PowerPoint presentation. 3 slides.

Additional Papers Cited by me or the Panel

Dunn A; RICC Francis; IJ.Doonan.(2002). Comparison of the Chapman–Robson and regression estimators of Z from catch-curve data when non-sampling stochastic error is present. *Fisheries Research 59*: 149–159.

McKenzie A (2016). Assessment of hoki (*Macruronus novaezelandiae*) in 2015. *New Zealand Fisheries Assessment Report 2016/01*. 88 p.

Millar RB (2014). A better estimator of mortality rate from age-frequency data. *Canadian Journal of Fisheries and Aquatic Sciences 72*:364–375.

Needle CL (2015). Using self-testing to validate the SURBAR survey-based assessment model. *Fisheries Research 171*: 78–86. DOI: 10.1016/j.fishres.2015.03.001.

Wilberg MJ; ME Livings; JS Barkman; BT Morris; JM Robinson (2011). Overfishing, disease, habitat loss, and potential extirpation of oysters in upper Chesapeake Bay. *Marine Ecology Progress Series 436*: 131–144.

Appendix 2: Copy of the Statement of Work for Martin Cryer

Statement of Work

National Oceanic and Atmospheric Administration (NOAA)
National Marine Fisheries Service (NMFS)
Center for Independent Experts (CIE) Program
External Independent Peer Review

61st Stock Assessment Workshop/Stock Assessment Review Committee (SAW/SARC)
Benchmark stock assessment for Atlantic surfclam

Background
The National Marine Fisheries Service (NMFS) is mandated by the Magnuson-Stevens Fishery Conservation and Management Act, Endangered Species Act, and Marine Mammal Protection Act to conserve, protect, and manage our nation's marine living resources based upon the best scientific information available (BSIA). NMFS science products, including scientific advice, are often controversial and may require timely scientific peer reviews that are strictly independent of all outside influences. A formal external process for independent expert reviews of the agency's scientific products and programs ensures their credibility. Therefore, external scientific peer reviews have been and continue to be essential to strengthening scientific quality assurance for fishery conservation and management actions.

Scientific peer review is defined as the organized review process where one or more qualified experts review scientific information to ensure quality and credibility. These expert(s) must conduct their peer review impartially, objectively, and without conflicts of interest. Each reviewer must also be independent from the development of the science, without influence from any position that the agency or constituent groups may have. Furthermore, the Office of Management and Budget (OMB), authorized by the Information Quality Act, requires all federal agencies to conduct peer reviews of highly influential and controversial science before dissemination, and that peer reviewers must be deemed qualified based on the OMB Peer Review Bulletin standards. (http://www.cio.noaa.gov/services_programs/pdfs/OMB_Peer_Review_Bulletin_m05-03.pdf). Further information may be obtained from www.ciereviews.org.

Scope
The Northeast Regional Stock Assessment Review Committee (SARC) meeting is a formal, multiple-day meeting of stock assessment experts who serve as a panel to peer-review tabled stock assessments and models. The SARC peer review is the cornerstone of the Northeast Stock Assessment Workshop (SAW) process, which includes assessment development and report preparation (which is done by SAW Working Groups or ASMFC technical committees), assessment peer review (by the SARC), public presentations, and document publication. This review determines whether or not the scientific assessments are adequate to serve as a basis for developing fishery management advice. Results provide the scientific basis for fisheries within the jurisdiction of NOAA's Greater Atlantic Regional Fisheries Office (GARFO).

The purpose of this meeting will be to provide an external peer review of a benchmark stock assessment for **Atlantic surfclam** *(Spisula solidissima)*. The requirements for the peer review follow. This Statement of Work (SOW) also includes Appendix 1: TORs for the stock assessment, which are the responsibility of the analysts; Appendix 2: a draft meeting agenda; Appendix 3: Individual Independent Review Report Requirements; and Appendix 4: SARC Summary Report Requirements.

Requirements
NMFS requires three reviewers under this contract (i.e. subject to CIE standards for reviewers) to participate in the panel review. The SARC chair, who is in addition to the three reviewers, will be provided by either the New England or Mid-Atlantic Fishery Management Council's Science and Statistical Committee; although the SARC chair will be participating in this review, the chair's participation (i.e. labor and travel) is not covered by this contract.

Each reviewer will write an individual review report in accordance with the SOW, OMB Guidelines, and the TORs below. All TORs must be addressed in each reviewer's report. No more than one of the

reviewers selected for this review is permitted to have served on a SARC panel that reviewed this same species in the past. The reviewers shall have working knowledge and recent experience in the application of modern fishery stock assessment models. Expertise should include forward projecting statistical catch-at-age models. Reviewers should also have experience in evaluating measures of model fit, identification, uncertainty, and forecasting. Reviewers should have experience in development of Biological Reference Points (BRPs) that includes an appreciation for the varying quality and quantity of data available to support estimation of BRPs. Knowledge of sedentary invertebrates, their fishery management and ecosystem issues would be useful.

Requirements for Reviewers

- Review the background materials and reports prior to the review meeting
- Attend and participate in the panel review meeting
    - The meeting will consist of presentations by NOAA and other scientists, stock assessment authors and others to facilitate the review, to provide any additional information required by the reviewers, and to answer any questions from reviewers
- Reviewers shall conduct an independent peer review in accordance with the requirements specified in this SOW and TORs, in adherence with the required formatting and content guidelines; reviewers are not required to reach a consensus.
- Each reviewer shall assist the SARC Chair with contributions to the SARC Summary Report Deliver individual Independent Review Reports to the Government according to the specified milestone dates
- This report should explain whether each stock assessment Term of Reference of the SAW was or was not completed successfully during the SARC meeting, using the criteria specified below in the "Requirements for SARC panel."
- If any existing Biological Reference Points (BRP) or their proxies are considered inappropriate, the Independent Report should include recommendations and justification for suitable alternatives. If such alternatives cannot be identified, then the report should indicate that the existing BRPs are the best available at this time.
- During the meeting, additional questions that were not in the Terms of Reference but that are directly related to the assessments may be raised. Comments on these questions should be included in a separate section at the end of the Independent Report produced by each reviewer.
- The Independent Report can also be used to provide greater detail than the SARC Summary Report on specific stock assessment Terms of Reference or on additional questions raised during the meeting.

Requirements for SARC panel

- During the SARC meeting, the panel is to determine whether each stock assessment Term of Reference (TOR) of the SAW was or was not completed successfully. To make this determination, panelists should consider whether the work provides a scientifically credible basis for developing fishery management advice. Criteria to consider include: whether the data were adequate and used properly, the analyses and models were carried out correctly, and the conclusions are correct/reasonable. If alternative assessment models and model assumptions are presented, evaluate their strengths and weaknesses and then recommend which, if any, scientific approach should be adopted. Where possible, the SARC chair shall identify or facilitate agreement among the reviewers for each stock assessment TOR of the SAW.
- If the panel rejects any of the current BRP or BRP proxies (for BMSY and FMSY and MSY), the panel should explain why those particular BRPs or proxies are not suitable, and the panel should recommend suitable alternatives. If such alternatives cannot be identified, then the panel should indicate that the existing BRPs or BRP proxies are the best available at this time.
- Each reviewer shall complete the tasks in accordance with the SOW and Schedule of Milestones and Deliverables below.

Requirements for SARC chair and reviewers combined:

Review both the Assessment Report and the draft Assessment Summary Report. The draft Assessment Summary Report is reviewed and edited to assure that it is consistent with the outcome of the peer review, particularly statements that address stock status and assessment uncertainty.

The SARC Chair, with the assistance from the reviewers, will write the SARC Summary Report. Each reviewer and the chair will discuss whether they hold similar views on each stock assessment Term of Reference and whether their opinions can be summarized into a single conclusion for all or only for some of the Terms of Reference of the SAW. For terms where a similar view can be reached, the SARC Summary Report will contain a summary of such opinions. In cases where multiple and/or differing views exist on a given Term of Reference, the SARC Summary Report will note that there is no agreement and will specify - in a summary manner – what the different opinions are and the reason(s) for the difference in opinions.
The chair's objective during this SARC Summary Report development process will be to identify or facilitate the finding of an agreement rather than forcing the panel to reach an agreement. The chair will take the lead in editing and completing this report. The chair may express the chair's opinion on each Term of Reference of the SAW, either as part of the group opinion, or as a separate minority opinion. The SARC Summary Report will not be submitted, reviewed, or approved by the Contractor.

If any existing Biological Reference Points (BRP) or BRP proxies are considered inappropriate, the SARC Summary Report should include recommendations and justification for suitable alternatives. If such alternatives cannot be identified, then the report should indicate that the existing BRP proxies are the best available at this time.

Foreign National Security Clearance
When reviewers participate during a panel review meeting at a government facility, the NMFS Project Contact is responsible for obtaining the Foreign National Security Clearance approval for reviewers who are non-US citizens. For this reason, the reviewers shall provide requested information (e.g., first and last name, contact information, gender, birth date, country of birth, country of citizenship, country of permanent residence, country of current residence, dual citizenship (yes, no), passport number, country of passport, travel dates.) to the NEFSC SAW Chair for the purpose of their security clearance, and this information shall be submitted at least 30 days before the peer review in accordance with the NOAA Deemed Export Technology Control Program NAO 207-12 regulations available at the Deemed Exports NAO website: http://deemedexports.noaa.gov/ and http://deemedexports.noaa.gov/compliance_access_control_procedures/noaa-foreign-national-registration-system.html. The contractor is required to use all appropriate methods to safeguard Personally Identifiable Information (PII).

Place of Performance
The place of performance shall be at the contractor's facilities, and at the Northeast Fisheries Science Center in Woods Hole, Massachusetts.

Period of Performance
The period of performance shall be from the time of award through August 31, 2016. Each reviewer's duties shall not exceed 12 days to complete all required tasks.

Schedule of Milestones and Deliverables: The contractor shall complete the tasks and deliverables in accordance with the following schedule.

No later than June 13, 2016	Contractor sends reviewer contact information to the COR, who then sends this to the NMFS Project Contact
No later than July 5, 2016	NMFS Project Contact will provide reviewers the pre-review documents
July 19-21, 2016	Each reviewer participates and conducts an independent peer review during the panel review meeting in Woods Hole, MA
July 21, 2016	SARC Chair and reviewers work at drafting reports during meeting at Woods Hole, MA, USA
August 4, 2016	Reviewers submit draft independent peer review reports to the contractor's technical team for review
August 4, 2016	Draft of SARC Summary Report, reviewed by all reviewers, due to the SARC Chair *
August 11, 2016	SARC Chair sends Final SARC Summary Report, approved by reviewers, to NMFS Project contact (i.e., SAW Chairman)
August 18, 2016	Contractor submits independent peer review reports to the COR and technical point of contact (POC)
August 25, 2016	The COR and/or technical POC distributes the final reports to the NMFS Project Contact and regional Center Director

* The SARC Summary Report will not be submitted to, reviewed, or approved by the Contractor.

Applicable Performance Standards
The acceptance of the contract deliverables shall be based on three performance standards: (1) The reports shall be completed in accordance with the required formatting and content; (2) The reports shall address each TOR as specified; (3) The reports shall be delivered as specified in the schedule of milestones and deliverables.

Travel
All travel expenses shall be reimbursable in accordance with Federal Travel Regulations (http://www.gsa.gov/portal/content/104790). International travel is authorized for this contract. Travel is not to exceed $23,000.

Restricted or Limited Use of Data
The contractors may be required to sign and adhere to a non-disclosure agreement.

Project Contacts
Dr. James Weinberg, NEFSC SAW Chair
Northeast Fisheries Science Center
166 Water Street, Woods Hole, MA 02543
James.Weinberg@noaa.gov Phone: 508-495-2352

Dr. William Karp, NEFSC Science Director
Northeast Fisheries Science Center
166 Water St., Woods Hole, MA 02543
william.karp@noaa.gov Phone: 508-495-2233

Allen Shimada, Technical Point of Contact
NMFS Office of Science and Technology
1315 East West Hwy, SSMC3, F/ST4, Silver Spring, MD 20910
allen.shimada@noaa.gov Phone: 301-427-8174

Patty Zielinski, COR
NMFS Office of Science and Technology
1315 East West Hwy, F/ST1, Silver Spring MD 20910
patty.zielinski@noaa.gov Phone: 301-427-8142

SoW Appendix 1. Terms of Reference for the SAW Working Group (61st SAW/SARC Stock Assessment)

The SARC Review Panel shall assess whether or not the SAW Working Group has reasonably and satisfactorily completed the following actions.

A. Atlantic surfclams

1. Estimate catch from all sources including landings and discards. Map the spatial and temporal distribution of landings, discards, fishing effort, and gross revenue, as appropriate. Characterize the uncertainty in these sources of data.
2. Present the survey data being used in the assessment (e.g., indices of relative or absolute abundance, recruitment, state surveys, age-length data, etc.). Use logbook data to investigate regional changes in LPUE, catch and effort. Characterize the uncertainty and any bias in these sources of data. Evaluate the spatial coverage, precision, and accuracy of the new clam survey.
3. Determine the extent and relative quality of benthic habitat for surfclams in the Georges Bank ecosystem to refine estimates of stock size based on swept area calculations.
4. Quantify changes in the depth distribution of surfclams over time. Review changes over time in surfclam biological parameters such as length, width, and growth.
5. Estimate annual fishing mortality, recruitment and stock biomass (both total and spawning stock) for the time series (integrating results from TOR 3, as appropriate) and estimate their uncertainty. Include a historical retrospective analysis to allow a comparison with previous assessment results and previous projections.
6. State the existing stock status definitions for “overfished” and “overfishing”. Then update or redefine biological reference points (BRPs; point estimates or proxies for BMSY, BTHRESHOLD, FMSY and MSY) and provide estimates of their uncertainty. If analytic model-based estimates are unavailable, consider recommending alternative measurable proxies for BRPs. Comment on the scientific adequacy of existing BRPs and the “new” (i.e., updated, redefined, or alternative) BRPs, particularly as they relate to stock assumptions.
7. Evaluate stock status with respect to the existing model (from previous peer reviewed accepted assessment) and with respect to any new model or models developed for this peer review.
    a. When working with the existing model, update it with new data and evaluate stock status (overfished and overfishing) with respect to the existing BRP estimates.
    b. Then use the newly proposed model and evaluate stock status with respect to “new” BRPs and their estimates (from TOR-5).
8. Develop approaches and apply them to conduct stock projections.
    a. Provide numerical annual projections (five years) and the statistical distribution (e.g., probability density function) of the OFL (overfishing level) (see Appendix to the SAW TORs). Consider cases using nominal as well as potential levels of uncertainty in the model. Each projection should estimate and report annual probabilities of exceeding threshold BRPs for F, and probabilities of falling below threshold BRPs for biomass. Use a sensitivity analysis approach in which a range of assumptions about the most important uncertainties in the assessment are considered (e.g., terminal year abundance, variability in recruitment).
    b. Comment on which projections seem most realistic. Consider the major uncertainties in the assessment as well as sensitivity of the projections to various assumptions.
    c. Describe this stock’s vulnerability (see “Appendix to the SAW TORs”) to becoming overfished, and how this could affect the choice of ABC.
9. Evaluate the validity of the current stock definition. Determine whether current stock definitions may mask reductions in sustainable catch on regional spatial scales. Make a recommendation about whether there is a need to modify the current stock definition.
10. Review, evaluate and report on the status of the SARC and Working Group research recommendations listed in most recent SARC reviewed assessment and review panel reports. Identify new research recommendations.

Clarification of Terms
used in the SAW/SARC Terms of Reference

On "Acceptable Biological Catch" (DOC Nat. Stand. Guidel. Fed. Reg., v. 74, no. 11, 1-16-2009):
Acceptable biological catch (ABC) is a level of a stock or stock complex's annual catch that accounts for the scientific uncertainty in the estimate of Overfishing Limit (OFL) and any other scientific uncertainty…" *(p. 3208) [In other words, OFL ≥ ABC.]*
ABC for overfished stocks. For overfished stocks and stock complexes, a rebuilding ABC must be set to reflect the annual catch that is consistent with the schedule of fishing mortality rates in the rebuilding plan. *(p. 3209)*
NMFS expects that in most cases ABC will be reduced from OFL to reduce the probability that overfishing might occur in a year. (p. 3180)
ABC refers to a level of ''catch'' that is ''acceptable'' given the ''biological'' characteristics of the stock or stock complex. As such, Optimal Yield (OY) does not equate with ABC. The specification of OY is required to consider a variety of factors, including social and economic factors, and the protection of marine ecosystems, which are not part of the ABC concept. (p. 3189)

On "Vulnerability" (DOC Natl. Stand. Guidelines. Fed. Reg., v. 74, no. 11, 1-16-2009):
"Vulnerability. A stock's vulnerability is a combination of its productivity, which depends upon its life history characteristics, and its susceptibility to the fishery. Productivity refers to the capacity of the stock to produce Maximum Sustainable Yield (MSY) and to recover if the population is depleted, and susceptibility is the potential for the stock to be impacted by the fishery, which includes direct captures, as well as indirect impacts to the fishery (e.g., loss of habitat quality)." (p. 3205)

Participation among members of a Stock Assessment Working Group:
Anyone participating in SAW meetings that will be running or presenting results from an assessment model is expected to supply the source code, a compiled executable, an input file with the proposed configuration, and a detailed model description in advance of the model meeting. Source code for NOAA Toolbox programs is available on request. These measures allow transparency and a fair evaluation of differences that emerge between models.

SoW Appendix 2. Draft Review Meeting Agenda

{Final Meeting agenda was provided at time of award}

61st Stock Assessment Workshop/Stock Assessment Review Committee (SAW/SARC): Benchmark stock assessment for A. Atlantic surfclam

July 19-21, 2016
Stephen H. Clark Conference Room – Northeast Fisheries Science Center
Woods Hole, Massachusetts

AGENDA* (version: Dec. 31, 2015)
TOPIC PRESENTER(S) SARC LEADER RAPPORTEUR

Tuesday, July 19
10 – 10:30 AM
Welcome **James Weinberg**, SAW Chair
Introduction **TBD,** SARC Chair
Agenda
Conduct of Meeting
10:30 – 12:30 PM Assessment Presentation (A. Surfclam) **Dan Hennen TBD**
12:30 – 1:30 PM Lunch
1:30 – 3:30 PM Assesssment Presentation (A. Surfclam) **Dan Hennen TBD**
3:30 – 3:45 PM Break
3:45 – 5:45 PM SARC Discussion w/ Presenters (A. Surfclam) **TBD,** SARC Chair **TBD**
5:45 – 6 PM Public Comments
7 PM (Social Gathering**)**

Wednesday, July 20
9:00 – 10:45 Revisit with Presenters (A. Surfclam) **TBD,** SARC Chair **TBD**
10:45 - 11 Break
11 – 11:45 Revisit with Presenters (A. Surfclam) **TBD,** SARC Chair **TBD**
11:45 – Noon Public Comments
12 – 1:15 PM Lunch
1:15 – 4 Review/Edit Assessment Summary Report (A. Surfclam) **TBD,** SARC Chair **TBD**
4 – 4:15 PM Break
4:15 – 5:00 PM SARC Report writing

Thursday, July 21
9:00 AM – 5:00 PM SARC Report writing

*All times are approximate, and may be changed at the discretion of the SARC chair. The meeting is open to the public. During “SARC Report writing”, on July 20 and 21, the public should not engage in discussion with the SARC.

SoW Appendix 3. Individual Independent Peer Review Report Requirements

1. The independent peer review report shall be prefaced with an Executive Summary providing a concise summary of whether they accept or reject the work that they reviewed, with an explanation of their decision (strengths, weaknesses of the analyses, etc.).
2. The report must contain a background section, description of the individual reviewers' roles in the review activities, summary of findings for each TOR in which the weaknesses and strengths are described, and conclusions and recommendations in accordance with the TORs. The independent report shall be an independent peer review, and shall not simply repeat the contents of the SARC Summary Report.
    a. Reviewers should describe in their own words the review activities completed during the panel review meeting, including a concise summary of whether they accept or reject the work that they reviewed, and explain their decisions (strengths, weaknesses of the analyses, etc.), conclusions, and recommendations.
    b. Reviewers should discuss their independent views on each TOR even if these were consistent with those of other panelists, but especially where there were divergent views.
    c. Reviewers should elaborate on any points raised in the SARC Summary Report that they believe might require further clarification.
    d. The report may include recommendations on how to improve future assessments.
3. The report shall include the following appendices:
    Appendix 1: Bibliography of materials provided for review
    Appendix 2: A copy of this Statement of Work
    Appendix 3: Panel membership or other pertinent information from the panel review meeting.

SoW Appendix 4. SARC Summary Report Requirements

1. The main body of the report shall consist of an introduction prepared by the SARC chair that will include the background and a review of activities and comments on the appropriateness of the process in reaching the goals of the SARC. Following the introduction, for each assessment reviewed, the report should address whether or not each Term of Reference of the SAW Working Group was completed successfully. For each Term of Reference, the SARC Summary Report should state why that Term of Reference was or was not completed successfully.

   To make this determination, the SARC chair and reviewers should consider whether or not the work provides a scientifically credible basis for developing fishery management advice. If the reviewers and SARC chair do not reach an agreement on a Term of Reference, the report should explain why. It is permissible to express majority as well as minority opinions.

   The report may include recommendations on how to improve future assessments.

1. If any existing Biological Reference Points (BRPs) or BRP proxies are considered inappropriate, include recommendations and justification for alternatives. If such alternatives cannot be identified, then indicate that the existing BRPs or BRP proxies are the best available at this time.

2. The report shall also include the bibliography of all materials provided during the SAW, and relevant papers cited in the SARC Summary Report, along with a copy of the CIE Statement of Work.

   The report shall also include as a separate appendix the assessment Terms of Reference used for the SAW, including any changes to the Terms of Reference or specific topics/issues directly related to the assessments and requiring Panel advice.

Appendix 3: Panel Membership and other pertinent information.

The review panel comprised Mike Bell, Coby Needle, and Martin Cryer, and was chaired by Mike Wilberg. Also present at the review table were Russ Brown, Jim Weinberg, Dan Hennen (lead assessment scientist, surf clams), Larry Jacobson (Chair, Invertebrate Subcommittee), and rapporteurs. A full list of participants who participated or attended at some time during the meeting follows:

Participant	Affiliation	Email
Michael Wilberg	University of Maryland - CES	wilberg@umces.edu
Coby Needle	Marine Scotland – Marine Lab-Aberdeen	c.needle@marlab.ac.uk
Mike Bell	Heriot-Watt University –Centre for Island Tech	m.c.bell@hw.ac.uk
Martin Cryer	Ministry for Primary Industries, Wellington	martin.cryer@mpi.govt.nz
Russ Brown	NEFSC	russell.brown@noaa.gov
Jim Weinberg	NEFSC	james.weinberg@noaa.gov
Larry Jacobson	NEFSC	larry.jacobson@noaa.gov
Dan Hennen	NEFSC	daniel.hennen@noaa.gov
Jessica Coakley	MAFMC	jcoakley@mafmc.org
Mike Simpkins	NEFSC	michael.simpkins@noaa.gov
Chris Legault	NEFSC	chris.legault@noaa.gov
Sheena Steiner	NEFSC	sheena.steiner@noaa.gov
Michele Traver	NEFSC	michele.traver@noaa.gov
Toni Chute	NEFSC	toni.chute@noaa.gov
Mark Terceiro	NEFSC	mark.terceiro@noaa.gov
Tom Hoff	Wallace & Associates	tbhoff@verizon.net
Daphne Munroe	Rutgers University	dmunroe@hsrl.rutgers.edu
Tom Alspach	Sea Watch International	talspach@goeaston.net
Eric Powell	University of Southern Mississippi	eric.n.powell@usm.edu
Bob Glenn	Mass. Division of Marine Fisheries	robert.glenn@state.ma.us
D.H. Wallace	Wallace & Associates	dhwallace@aol.com
Doug Potts	NMFS/GARFO	douglas.potts@noaa.gov
Wendy Gabriel	NEFSC	wendy.gabriel@noaa.gov
Jon Deroba	NEFSC	jonathan.deroba@noaa.gov
Charles Perretti	NEFSC	charles.perretti@noaa.gov